UCI-ITS-AS-WP-02-12

The Design and Implementation of an On-Line Travel and Activity Survey

UCI-ITS-AS-WP-02-12

James E. Marca

Department of Civil and Environmental Engineering and
Institute of Transportation Studies
University of California, Irvine, jmarca@uci.edu

August 2002

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

THE DESIGN AND IMPLEMENTATION OF AN ON-LINE TRAVEL AND ACTIVITY SURVEY

James E. Marca[1]

ABSTRACT

Long records of activities and travel for individuals, essential for understanding the dynamic changes in traveler behavior, do not exist due to the difficulty of collecting such data. To address this need, an on-line activity survey was designed that is tightly intertwined with real-time position data streaming over wireless data links from in-vehicle GPS data collection devices. While the technology to construct such a survey has existed for some time, the author has been unable to find other published examples of such a survey system. Some preliminary observations of the system based on a small, informal pilot survey are reported.

Keywords: activity survey, on-line, GPS, wireless

INTRODUCTION

This paper presents the design of a web-based activity survey. This survey uses a wireless and GPS enabled extensible data collection unit (EDCU), documented fully in a companion working paper by the author (Marca 2002b). The next section reviews some recent applications of GPS technology to travel surveys. Following that is a description of a new survey instrument which integrates the real-time EDCU data into an activity and travel survey instrument. Although the enabling technologies have existed for at least 3 years, the author is not aware of any other implementation of a dynamic web-based activity survey that integrates a real-time stream of GPS data. Section 5 presents some preliminary implementation lessons gleaned from a very small pilot application. The results identify some areas that need improvement. Further, the pilot survey indicated that the most likely usage scenario will be one that asks the respondent to complete the web survey, and then follows up the initial survey with a reminder call and a series of web pages tailored to ask about the most important missing activities.

COLLECTING ACTIVITY DATA

[1]Institute of Transportation Studies, University of California, Irvine. E-mail: jmarca@translab.its.uci.edu

Activity models demand a much richer data set than that required for the traditional trip-based travel planning process. An excellent example is contained in the ALBATROSS book (Arentze and Timmermans 2000), which presents a detailed activity model framework alongside an extensive discussion of the model's data needs and the survey diary that they developed to collect that data. While details concerning each activity and the person's motivations are important, equally important is to understand the long-term dynamics of the person's travel and activity patterns. Three technological enhancements to travel surveys have been developed recently which all contribute to this goal. First is the computer aided survey instrument. Second is the GPS data collection device. Third is the ability to transmit data wirelessly, in near real-time. The contributions of each of these are discussed in the following paragraphs.

Computers accurately record and remember facts. This has been used to design surveys that can prompt a user to fix holes and inconsistencies in their responses. Kalfs and Saris (1997) give a thorough accounting of the ways in which computer interview techniques improve upon pencil and paper time use diaries. The findings demonstrate that errors decline in the computer aided surveys, but that more can be done to reduce under-reported trips. They conclude their review of computer aided survey capabilities by recommending that more be done to automate the data collection process, and identify GPS technologies as a new and promising way to automate time and position recording of trips.

Murakami and Wagner (1999) show that a GPS data logger is excellent for eliminating eliminating missing trips and collecting accurate activity durations and start times, although the *in-vehicle* GPS device used here measures travel between parking spots, not travel to and from actual activity destinations. While GPS antennas have errors associated with them that must be handled appropriately (Wolf et al. 1999), these errors are largely mechanical and quantifiable. This is a tremendous improvement over the variable and intractable human errors that plague travel diaries.

Wolf et al. (2001) explore using GPS devices as the *exclusive* means of collecting travel and activity data. They recommend relying upon a highly precise geographic information system (GIS) to link a destination coordinate with the underlying land use, and a likely activity. The difficulty with this approach is that it requires an expensive GIS system, and even so cannot guarantee that the activity at the destination that is estimated is in fact the actual activity. A follow-up interview is still required.

The approach of Bachu et al. (2001), which collects GPS measurements that are used to create memory-jogging maps for a follow-up activity survey, is very close to the approach taken here. The difference is that our GPS system takes advantage of the third technological innovation, wireless data communication. Rather than waiting until after the field study is over to collect the GPS device and generate an activity survey instrument, our system collects GPS data continuously by streaming that data from each EDCU device over a cellular digital packet data (CDPD) modem. All of the fielded EDCUs communicate with a central survey base station that logs the GPS record streams directly to a database. The integration of GPS data into a dynamically generated activity survey instrument is documented in the next section.

WEB-BASED, EDCU-ENHANCED ACTIVITY SURVEY

Activity survey goals

The purpose of the activity survey is to gather qualitative information about destinations and travel from the respondent, assuming the location and time of the day's travel paths and destinations are already known. The primary design criterion for the survey instrument is to be easy on the user. That coupled with the fact that the location and time are already known means that many of the detailed questions that are common in time use and activity surveys are not asked here. A first generation survey was developed around a simple map rendering engine coupled with an HTML form input asking for a description of the activity at the destination. This provided a strawman implementation which was refined piece by piece, as detailed in the following sections.

Displaying a map

The first aspect of the survey to be improved was the map. As a map is the visual and logical link between the collected GPS data and the survey, its handling is crucial to the effective implementation of the survey. An object oriented map generator, written in Perl, was developed to convert the raw GPS files in the database into images and aggregate measures of the travel.

The core classes of the map generator represent the map image to be displayed, the travel to the destination, and the activity at the destination. A day might produce several trips. To avoid confusion about the meaning of "trip" in this context, the word "trace" is used to represent a sequence of GPS records starting at one time and ending at another. The respondent selects the day or days to be viewed, and the program fetches the corresponding list of unique `traceids` from the database. Each `traceid` is used to create an instance of the trace image class. These objects grab their GPS points from the database and generate a plot of the trace layered on top of an appropriate map of the travel area. The first time a trace image object is created, the object calculates averages, durations, and other data-driven annotations for the trace. These are saved in a separate quantitative annotation table in the database, along with state information about the internals of the trace image object. This state information allows subsequent object instantiation and web page rendering to be much faster.

The plots are static images, but based on the preliminary trials, the images are rendered at three scales as appropriate for the length of the travel. The shortest traces are plotted on a neighborhood scale, in which features like individual blocks are readily apparent. Medium length traces are plotted on a scale at which local streets appear to be a dense mesh, while any traces longer than the extent of the medium map are drawn on an enormous map scale which can fit most of the Los Angeles metropolitan area. Maps are rendered as mosaics when necessary. Traces can span 3 or 4 map squares at one scale, before being bumped up to the next larger scale. This mechanism avoids having to implement a panning and zooming interface, while still maintaining a degree of detail that is appropriate to the trace in question. A survey screen-shot, with a map drawn at the neighborhood scale, is shown in figure 1.

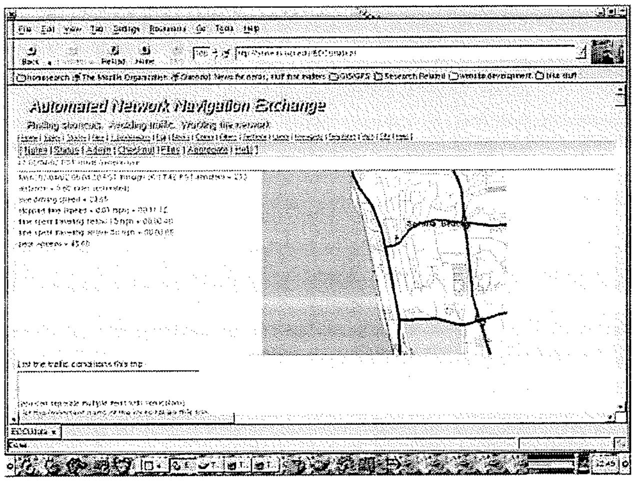

FIG. 1. The initial screen of a survey. Other trips are visible by scrolling down. The web version renders the travel path in red, yellow, or green, depending upon the travel speed, on a grayscale map.

Travel and activity annotations

The strawman HTML text form was expanded into separate forms for the travel conditions, the route, the destination name, the activity at the destination, and the people involved in the activity. A screenshot sequence similar that which would be seen by a first-time user is shown in figure 1 and figure 2. The HTML page is generated piece by piece for each trace being shown. First the `traceid` is used to create a trace image object, which is used to generate a map of the travel as explained above. Then other classes are instantiated as needed to fetch the numerical annotations, as well as any past survey entries of each trace. These objects all pass data to an HTML template, which renders the data in a readable layout. This separation of content from design is a highly recommended practice for data driven web sites.

To the left of the map in figure 1 are the numerical annotations, dates, and times that characterize the travel. Below the map and its annotations are the survey questions, shown in figure 2. This pattern is repeated for each separate trace. If the respondent wishes to complete the survey for several days at once, each day is listed in order, with a blue bar and bold text highlighting the day breaks. All of these formatting aspects are easily manipulated by editing the template without having to change the content-generating code. When information is entered into the forms, the response is saved to the database and used to create checkboxes for future entries. An example of a survey form after several weeks of data entry is shown in figure 3. Each list of options has been sorted in space and in time according to the closest past observation relative to the current location. This was found to work very well to put the most likely option at the top of this list. It should also be noted that the respondent can check or enter as many options as apply.

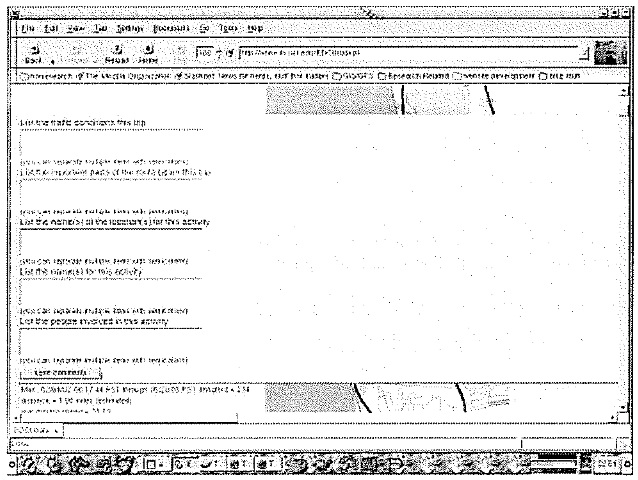

FIG. 2. Text forms for the travel conditions, the route, the destination name, the activity at the destination, and the people involved in the activity. This user has not yet entered any information.

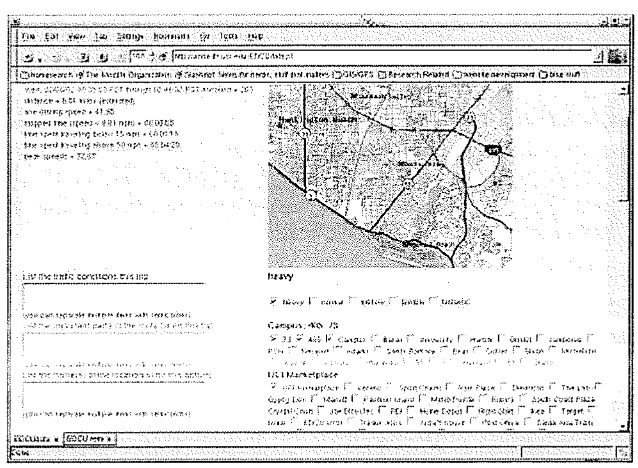

FIG. 3. The survey dialog with several weeks of travel options entered and available as checkboxes. The map shown is rendered at the middle scale.

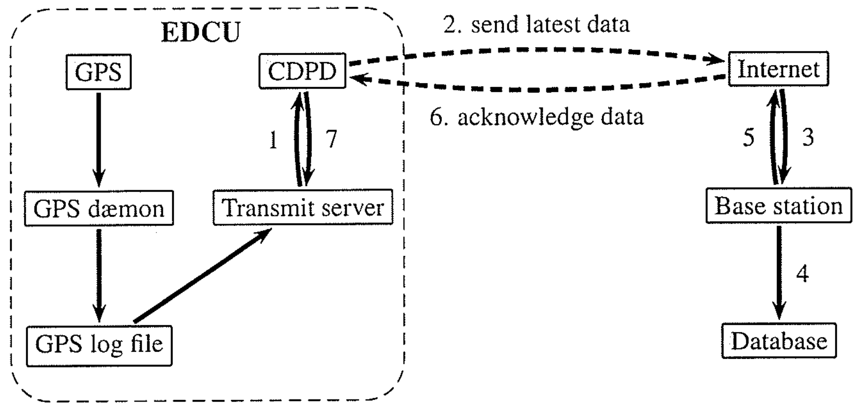

FIG. 4. Flow of information from the GPS to the database. If the CDPD link fails to deliver message 2 or message 6, then the transmit server does not receive acknowledgement that the data has been stored to the base station database.

SURVEY FIELD TEST AND LESSONS LEARNED

The survey system has been installed and tested using four volunteers in an informal pilot survey. These volunteers placed an EDCU in one of their vehicles, and were asked to annotate their activities and destinations. Invariably the respondents stopped annotating their activities after a short period of time, but they all continued to use the EDCUs in their vehicles for much longer periods. The main result of the pilot survey was to gather important information on what can be improved in the system, as reported in the following subsections.

Detecting a stop

The operational logic of the wireless EDCU data collection process is shown in figure 4. There are two sources of errors with the way this process was implemented for the pilot survey. First, the CDPD link is not always reliable. Referring to the numbered messages in figure 4, if the CDPD link fails for message 2 or 6, then the transmit server will not know if the base station received the data. This has led to the design of a forking Transmit server in the EDCU. A child process is spawned to send the data, and it dies only upon acknowledgment of receipt by the base station. The parent process periodically collects all persistent children and resends their messages in a new child process.

Second, the base station has no way of knowing whether a mobile unit is stopped and powered down, or collecting data but out of CDPD coverage range. The mobile units were programmed to send a "shut down" message when the units turned off, but this strategy turned out to be unsuccessful, as the battery backup uninterruptible power supply (UPS) often did not have enough power to gracefully shut down. Another problem with this strategy arises for vehicles which to not power down their cigarette lighters when the ignition is switched off. In these vehicles, the EDCUs will continue to transmit (stationary) points until the internal motion sensor shuts down the unit.

The base station was reprogrammed to attempt to automatically detect a stop. The algorithm was only partially successful, with a bias towards adding too many stops rather than too few. Finally, if the CDPD connection is weak and/or a trip takes less than a few minutes, the system may entirely miss a trip and a stop. Note that the data still resides on the EDCU's internal log in these cases, and the transmit server could be extended to send these short trips at some other time.

User interface

The user interface for the on-line survey has both strengths and weaknesses. The interface is easy to use at first, and presents information clearly. However, aside from very short trips, the automatically generated maps are usually mosaics of two or more maps. This means that they extend beyond the design map size of 400×300 pixels. While the maps are quite clear and easy to read, the respondent must scroll right in order to see them, and then scroll back left to enter the survey information. This may cause confusion for users who are unaccustomed to information that extends beyond the boundaries of a browser window. The solution to this problem is not obvious, as it is endemic throughout the Internet whenever a web page does not fit easily within the lowest common denominator 640×480 pixel screen. Frames are often used to address this problem, but they bring their own cross-browser compatibility issues.

A more serious problem is the potentially infinite growth of the checkbox lists of past entries, as shown in figure 3. As more and more entries are recorded, the lists begin to dominate the survey form, and become more of a hindrance than an aid. The obvious solution is to limit the numbers of checkboxes that are shown, with a bit of Javascript embedded to enable an option to see more choices if desired. After an analysis of the collected pilot data, the current distance and time based sort has been replaced by a list of most likely activities. This list is determined by estimating a kriging surface (Cressie 1993) for the activity frequency at each destination point. This technique is beyond the scope of this paper, and is documented in a companion paper by the author (Marca 2002a).

CONCLUSION

This paper has presented a new activity and travel survey instrument. The survey dynamically incorporates streaming GPS data, as well as the respondent's own past responses. This implementation demonstrates that the combination of a real-time GPS stream and an activity survey is practical and feasible. A small pilot survey was conducted to test the system, and it was found that the activity survey still needs several improvements. The main usability barrier was the growing list of past activity entries. This has been shortened to only the most likely activities, by using a kriging surface that is documented fully in a working paper by the author (Marca 2002a). It was also found that the wireless connection often caused gaps in recorded travel, or missed very short trips entirely.

The best use of this survey is likely to be in the extended data collection application as intended. However, since the respondents demonstrated a rather quick onset of survey fatigue, a practical implementation will most likely require a follow up reminder call. One advantage of collecting data continuously and rendering surveys dynamically is that the reminder call can be followed by setting the web site to generate first survey pages that pertain to the most interesting and/or least reported travel. Thus the waning interest of the respondents can be used to gather data that will add the most information content.

REFERENCES

Arentze, T. and Timmermans, H. (2000). *ALBATROSS: A Learning Based Transportation Oriented Simulation System.* CIP-DATA Koninklijke Bibliotheek, Den Haag, Netherlands.

Bachu, P. K., Dudala, T., and Kothuri, S. M. (2001). "Prompted recall in global positioning system survey: Proof-of-concept study." *Transportation Research Record*, (1768), 106–113.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* John Wiley & Sons, Inc., New York, revised edition.

Kalfs, N. and Saris, W. E. (1997). "New Data Collection Methods in Travel Surveys." *Activity-Based Approaches to Travel Analysis*, D. Ettema and H. Timmermans, eds., Elsevier Science Inc., Tarrytown, NY, USA, chapter 13, 243–261.

Marca, J. E. (2002a). "Collecting activity data from gps readings. Working paper.

Marca, J. E. (2002b). "The tracer data collection system: Implementation and operational experience. Working paper.

Murakami, E. and Wagner, D. P. (1999). "Can using global positioning system (GPS) improve trip reporting?." *Transportation Research C*, 7(2/3), 149–165.

Wolf, J., Guensler, R., and Bachman, W. (2001). "Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data." *Transportation Research Record*, (1768), 125–134.

Wolf, J., Hallmark, S., Oliveira, M., Guensler, R., and Sarasua, W. (1999). "Accuracy issues with route choice data collection by using global positioning system." *Transportation Research Record*, (1660), 66–74.